Investigating the Factors Affecting Risky Levels of Alcohol Consumption among Students Using Machine Learning Approach

Sara Fariha Shanchary
North South University
Dhaka, Bangladesh

Md Naved Meraz
North South University
Dhaka, Bangladesh
naved.meraz@northsouth.edu

Ayman Ibne Hakim
North South University
Dhaka, Bangladesh

Chowdhury Nafis Faiyaz
North South University
Dhaka, Bangladesh

Muhammad Shafayat Oshman
North South University
Dhaka, Dhaka, Bangladesh

ABSTRACT

Addressing the pressing issue of alcohol consumption among students is crucial for the well-being of young individuals and the community. Understanding and mitigating alcohol related challenges faced by young people is essential for their healthy development. This study utilizes machine learning models to analyze data from a Portuguese school, aiming to link students' alcohol consumption levels to personal and familial factors. The primary objective is to identify key factors associated with high alcohol consumption among students. After data preprocessing on their dataset, we employed various machine learning algorithms, including hyperparameter-optimized Decision Tree, Random Forest, Boosting, and Ensemble Learning. Our findings revealed that the decision tree algorithm performed well in predicting risky alcohol consumption for our target research question. Our selected feature subset showed strong positive correlation with the target variable, achieving an accuracy of 80.9 percent on the test set and 98.43 percent on the train set. Notably, this project breaks new ground by using explainable artificial intelligence to add reason to our prediction based on students' familial relationships, expanding upon previous research that also focused on demographic factors.

CCS CONCEPTS

• **Computing methodologies** → Artificial intelligence; *Ensemble methods*; *Feature selection*; **Supervised learning**.

KEYWORDS

alcohol consumption, alcohol use disorder, machine learning, family relationships

ACM Reference Format:

Sara Fariha Shanchary, Md Naved Meraz, Ayman Ibne Hakim, Chowdhury Nafis Faiyaz, and Muhammad Shafayat Oshman. 2024. Investigating the Factors Affecting Risky Levels of Alcohol Consumption among Students Using Machine Learning Approach. In *Proceedings of 2024 6th Asia Conference on Machine Learning and Computing (ACMLC '24).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The consumption of alcohol among students is a prevalent and well-documented phenomenon that can have grave consequences for both individuals and society as a whole. It has significant societal repercussions, including academic decline, violence, injuries, mental health issues, and fatalities [5]. Understanding the determinants of this risky behavior is vital, especially because young individuals are often introduced to alcohol early, leading to societal challenges [6] which can continue further into adult life. This is a probable pattern for many young individuals and thus is a societal issue. Unchecked alcohol consumption can develop into risky drinking patterns that can affect work, focus, and health [4]. Recovery from alcohol use disorder (AUD) often involves medications with serious side effects [15], making it imperative to investigate the underlying reasons for AUD. We employ machine learning to predict high alcohol consumption among school students in Portugal—focusing on the factor of unstable family relationships. To ensure transparency and mitigate biases, we utilize LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), model-agnostic explainable artificial intelligence (XAI) techniques. This study's contributions encompass:

- Utilizing multiple classifiers—Decision Tree, Random Forest, Soft voting, Hard voting, Out-of-bag error (OOB), extreme gradient boosting (XGBoost), Adaptive Boosting (AdaBoost).
- Performing hyperparameter optimization for the best-performing classifiers.
- Applying XAI methods (LIME and SHAP) to identify influential features.

The subsequent sections are organized as follows: Section II offers a critical literature review, followed by the research methodology in Section III. Section IV presents the experimental findings, while Section V discusses limitations. Finally, Section VI concludes by summarizing key findings and suggesting avenues for future research.

2 LITERATURE REVIEW

In previous approaches to identify factors contributing to AUD, individual substances have fallen short in capturing complex interrelationships, as highlighted by Ruberu et al.'s study [12] in 2022.

ACMLC '24, July 26–28, 2024, Bangkok, Thailand

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

Hence, the authors propose a joint model to predict hazardous substance use. Trained on individual, familial, and socio-demographic data of adolescent substance users, their model surpassed traditional regression models in accuracy. It identifies risk factors like early tobacco use, reduced parental attachment, and heightened early-life stress for finding at-risk adolescents. Limitations include cross-sectional design, inability to establish causality, and a small sample size for youths regularly using alcohol and cannabis; potentially limiting generalizability. Despite these constraints, the study provides valuable insights into adolescent hazardous substance abuse.

In a study by Singh et al. (2022) [14], machine learning and deep learning techniques were used to analyze alcohol consumption patterns among secondary school students in relation with their enrolled courses. The study explores various machine learning models with XGBoost and Random Forest exhibiting the highest accuracy levels. The study reveals a higher proportion of alcohol use among students in the Portuguese course compared to the Mathematics course. The same dataset [10] was previously used in [13] to identify alcohol addiction using a Decision Tree algorithm, achieving 93% accuracy. This study identifies 'going out with friends' as the most influential factor in alcohol consumption, while emphasizing that it lacks a validation set and is solely focused on the Decision Tree algorithm.

Shamala Palaniappan et al. [11] used neural networks to predict the alcohol consumption behaviors among higher secondary school students. Their study compares the prediction accuracies of two neural network algorithms: a self-tuning multi-layered perceptron classifier (AutoMLP) and the standard multi-layered perceptron (MLP), using the dataset from [10]. Both MLP and AutoMLP models constructed based on the training set achieved classification on the test set using 10–fold cross–validation. The study found that AutoMLP outperformed standard neural networks, achieving an accuracy of 64.54% compared to 61.78%.

A 2021 study, funded by Seoul St. Mary's Hospital [7], aimed to identify hazardous drinkers and assess the severity of alcohol-related issues using deep learning. The study, based on a large South Korean population survey, employed conventional machine learning algorithms (Logistic Regression, SVM, Random Forest, and K-Nearest Neighbors) alongside deep learning, consistently finding deep learning to outperform. The study unveils that hazardous drinkers had higher average intakes of energy, carbohydrates, fats, proteins, and dietary components. However, the dataset lacks crucial features like social and economic status, or family history of alcohol abuse; limiting a comprehensive understanding of alcohol-related problems.

Marcon et al. [9] surveyed online, 4840 Brazilian medical students, with 53.03% exhibiting high-risk alcohol consumption. Utilizing a dataset of 93 predictors spanning demographic, personal, medical student domain, and mental health-related variables, machine learning models (Lasso, Elastic-Net Regularized Generalized Linear Models, artificial neural network, and Random Forest) effectively differentiated high-risk from low-risk drinkers. The average Area Under Curve (AUC) scores in cross-validation were approximately 0.72 to 0.73. The study identified the top 30 relevant features associated with high-risk alcohol consumption, including factors like tobacco or cannabis use, monthly family income, relationship status, physical activity, and sexual orientation; influencing the model's classification of high-risk drinking.

Perceptions of peer behavior significantly influence teen drinking. In a study by [2], theory-driven covariate selection was combined with machine learning to reveal that adolescents often overestimate their peers' alcohol use and abuse. Using a sample of 14,738 respondents, Generalized Linear Models, Random Forest algorithms, and XGBoost Regression were used. XGBoost Regression achieved the highest accuracy (75.83%). The findings highlighted the positive associations between both–actual peer drinking and normative misperception with adolescent alcohol consumption, underscoring the significant role of peer norms. The study suggests that interventions targeting peer norms may effectively reduce adolescent alcohol consumption.

One of the first studies to develop and validate a machine learning model for predicting adolescent alcohol use in a cross-study, cross-cultural setting [1] used data from two groups of adolescents: one in Canada and one in Australia. The study used seven machine learning algorithms, creating models for different clusters. The elastic-net algorithm demonstrated superior performance, predicting alcohol consumption levels with over 85% accuracy in both group samples. Key predictors included externalizing psychopathologies, baseline alcohol use, and sensation-seeking personality profiles. However, the study's limitation to two countries restricts the generalizability of findings to other cultural settings.

3 METHODOLOGY

Our research's key objective is to identify multiple models that can predict with good accuracy the noticeable traits associated with heavy drinking among students. For our research and analysis, we have used the dataset from [10], that contains records of secondary school students in Portugal in two distinct courses: Mathematics and Portuguese language courses, along with their demographic and personal information. The block diagram for our research methodology is shown in Figure 1.

3.1 Block diagram

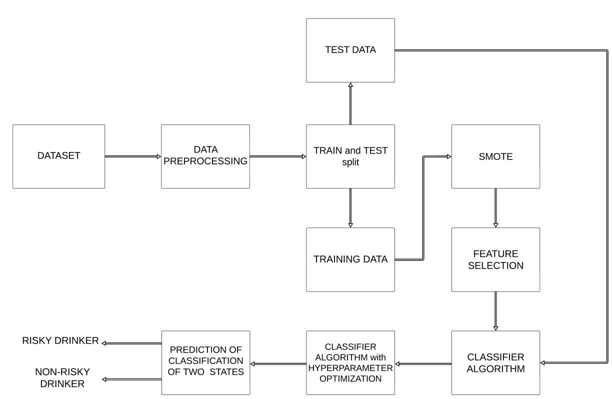

Figure 1: Block diagram of the research methodology

Table 1: Dataset Attributes

Attribute	Description(datatype)
age	student's age (numeric)
famsize	family size (binary)
Pstatus	parents' cohabitation status (binary)
Medu	mother's education (numeric)
Fedu	father's education (numeric)
Mjob	mother's job (nominal)
Fjob	father's job (nominal)
guardian	student's guardian (nominal)
famsup	family educational support (binary)
famrel	quality of family relationships (numeric)
alc	weekly alcohol consumption (numeric)

The preprocessing involved merging similar columns, converting categorical features into numeric using One Hot Encoder, and scaling all features with MinMaxScaler. We noticed the attribute size has increased to 45 after encoding. The dataset was split into an 80:20 train-test ratio. Synthetic Minority Oversampling Technique (SMOTE) was applied to the train set to address class imbalance. Feature selection was performed on the SMOTE dataset, and classifiers were applied to both the train and test sets. Hyperparameter optimization was conducted on the classifiers to enhance classification performance. Finally, predictions were made to categorize individuals as "risky drinker" (class 1) or "non-risky drinker" (class 0) based on classifier's output.

3.2 Dataset

The study began by merging data from Arithmetic and Portuguese language classes. Post-merge, we retained 11 relevant columns and discarded the rest as irrelevant. The aggregation of workday and weekend alcohol consumption columns formed a new variable, 'weekly alcohol consumption'. Alcohol consumption levels were measured on a scale of 1 to 5, denoting very low to very high consumption, respectively. Based on cumulative weekly intake, individuals surpassing level 4 were identified as unsafe drinkers. Key features for the analysis are detailed above in Table 1.

3.3 Research question

We assessed multiple classification algorithms on our manually curated subset of features to identify the best-performing model. By employing XAI, we sought insights into the predictions of this top-performing model. We hypothesized that alcohol consumption is influenced by a lack of family or educational support, aligning with existing research indicating the impact of parental and educational support systems on student alcohol consumption; actively involved parents correlate with lower chances of alcohol misuse in children. The target question for confirming the hypothesis is as follows:

- Does the quality of family relationships, parents' cohabitation status have an impact on alcohol consumption?

The classifiers used for the research question are discussed below.

3.4 Classifiers used

Decision Tree. Decision trees are a versatile supervised machine learning algorithm used for classification and regression. They create a tree-like model where internal nodes represent features and leaf nodes signify class labels or prediction values. The algorithm splits data at each internal node based on feature values, directing it to the appropriate child node. The root node is selected by maximizing information gain and minimizing entropy, as calculated using equation (1) and equation (2):

$$Entropy(p) = -\sum_{i=1}^{c} p_i(t)\log_2 p_i(t) \tag{1}$$

$$Gain = Entropy(p) - \sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \tag{2}$$

Here c = number of classes, k = number of partitions, n = number of parent nodes, and n_i = number of child nodes. The Gini index (3) is considered for each node and hyperparameter tuning is performed for multiple max depths.

$$Gini = 1 - \sum_{i=1}^{c} p_i(t)^2 \tag{3}$$

In the context of this research, the Gini impurity criterion was used for splitting nodes.

Random Forest. Random Forest, an ensemble learning approach, is used for classification, regression, and diverse tasks. It constructs multiple Decision trees during training, with each tree based on a bootstrap sample of the training data. At each node, a random subset of features is considered for splitting, mitigating overfitting and improving generalization. In this study, Random Forest was applied to predict risky alcohol consumption by using the entropy criterion for node splitting.

Soft Voting. Soft voting is an ensemble learning voting method that enhances prediction accuracy by combining the outputs of multiple machine learning models. It begins by averaging the predicted probabilities from each model for each class. The class with the highest average probability is then selected as the final ensemble prediction.

Hard Voting. Hard voting is an ensemble learning approach that combines the outputs of multiple individual models or classifiers to make a final prediction. Each model independently predicts the class label for a given input, and the class label with the most votes is selected as the ultimate prediction.

OOB Evaluation. OOB evaluation is commonly employed in ensemble learning, especially in bagging-based algorithms like random forests as it offers a performance estimate without requiring a distinct validation set. This technique predicts the target variable for each data point in the training set using all trees in the forest, excluding the one trained on that specific data point. The OOB error rate, calculated with (4), represents the proportion of data points for which the predicted target variable is incorrect.

$$OOB_(error) = \frac{\text{number of incorrect predictions}}{\text{total number of data points}} \tag{4}$$

Table 2: Accuracy for Train and Test set

Algorithms	Train accuracy	Test accuracy
Decision Tree	98.43%	80.90%
Random Forest	98.43%	77.50%
OOB evaluation	98.43%	77.50%
AdaBoost	98.43%	78.95%
XGBoost	96.64%	78.47%
Soft Voting	96.86%	79.43%
Hard Voting	96.86%	77.99%

XGBoost. XGBoost is a highly scalable and efficient gradient boosting implementation extensively applied in various machine learning tasks. It operates by sequentially constructing a series of weak decision trees, with each tree trained to minimize the loss of the preceding one. This iterative process continues until a predefined stopping criterion is satisfied. XGBoost employs gradient tree boosting, a more efficient approach to training tree models compared to traditional boosting methods. In this context, 100 parameters were utilized in XGBoost.

AdaBoost. AdaBoost is an ensemble machine learning algorithm that aggregates multiple weak learners to form a robust learner. It functions through iterative training of weak learners on weighted versions of the training data. The data point weights are adjusted after each iteration to emphasize misclassified points from the current weak learner. This cycle persists until a stopping condition is fulfilled, like reaching a maximum iteration limit or achieving a specific accuracy threshold. In this study, the decision tree was employed as the weak learner, with a learning rate set at 0.4.

4 RESULTS AND DISCUSSIONS

The train and test dataset contains 835 and 209 instances respectively and 11 attributes from original dataset were chosen for our research question.

Table 2 presents the accuracies of the top-performing classification algorithms following grid search on both train and test data. In addition to accuracy, the study assessed various performance metrics such as TPR (True Positive Rate), FPR (False Positive Rate), Precision (Positive Predictive Value), Recall (Sensitivity), and F-Measure (evaluation metric that measures a model's accuracy) to enable a comprehensive comparison of the utilized algorithms.

Table 3 and Table 4 show that in each case, Decision Tree and Random Forest algorithm performed the best among all other algorithms.

A receiver operating characteristic (ROC) curve is a graphical representation of the performance of a classification model. It plots the sensitivity (TPR) against the specificity (1 − FPR) for different cut-off values. The ROC curve for our model is shown in figure 2. The AUC is 0.89, which indicates that the model is performing well. The predictions were further analyzed using XAI which will be discussed below.

Table 3: Detailed evaluation on Train set

Algorithms	TP rate	FP rate	Precision	Recall	F1 score
Decision Tree	0.97	0.004	1.00	0.97	0.98
Random Forest	0.98	0.01	0.99	0.98	0.98
OOB Evaluation	0.98	0.01	0.99	0.98	0.98
AdaBoost	0.98	0.006	0.99	0.98	0.98
XGBoost	0.95	0.02	0.98	0.95	0.97
Soft voting	0.94	0.004	1.00	0.94	0.97
Hard voting	0.94	0.004	1.00	0.94	0.97

Table 4: Detailed evaluation on Test set

Algorithms	TP rate	FP rate	Precision	Recall	F1 score
Decision Tree	0.79	0.18	0.74	0.79	0.76
Random Forest	0.76	0.21	0.70	0.76	0.73
OOB Evaluation	0.76	0.21	0.70	0.76	0.73
AdaBoost	0.78	0.20	0.71	0.78	0.78
XGBoost	0.76	0.20	0.71	0.76	0.73
Soft voting	0.80	0.21	0.71	0.80	0.75
Hard voting	0.72	0.18	0.72	0.72	0.72

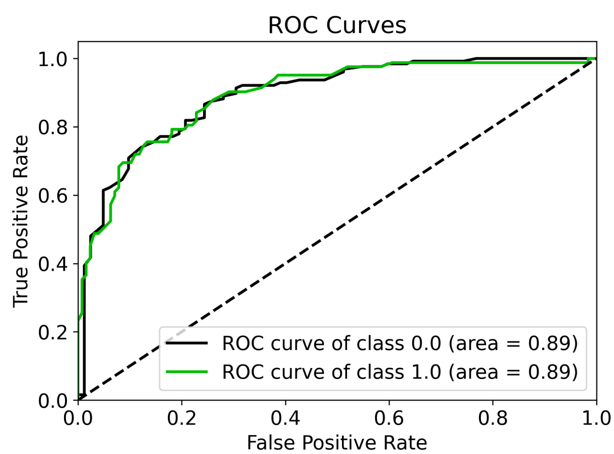

Figure 2: ROC curve on test set for Random Forest

4.1 Explainable Artificial Intelligence

With machine learning models becoming "black boxes", understanding their predictions has become challenging. To address this, there is a growing emphasis on machine learning interpretability and explainability, achieved through tools like XAI [3]. While many notable works predicted alcohol consumption using this dataset, we have employed two XAI tools—SHAP and LIME—to analyze prediction reasons. Feature values help identify these reasons, and the best grid of the random forest classifier was used for model fitting in both LIME and SHAP explanations.

LIME. LIME is a model-agnostic machine learning tool that helps us interpret ML models. LIME helps interpret reason for a specific prediction.

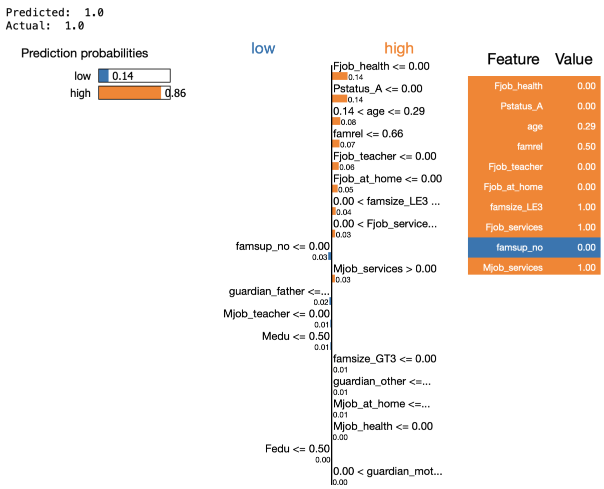

Figure 3: LIME visualization for prediction reason

In figure 3, the results of one specific prediction from test set obtained show that the attribute, Fjob health (father's job in health), holds the highest feature importance with a score of 14%. This is followed by Pstatus_A (parental status living apart) at 14% and famrel (quality of family relationships) at 7%. Notably, when the father's job is not health-related (Fjob health value 0.00 in the chart), it significantly contributes to the high label. Features highlighted in orange contribute to class 1, risky drinker, while those in blue impact class 0, non-risky drinker. Consequently, the cumulative feature importance scores lead to this prediction of high alcohol consumption. In summary, both parents working in service jobs, low family relationship quality, and parents living apart collectively led the model to predict this student as a high-risk drinker.

SHAP. SHAP is a mathematical method to explain the predictions of machine learning models. SHAP helps visualize across all samples.

To explain the prediction rationale, we used the SHAP explainer method. The beeswarm plot in figure 4 displays the contribution of the most critical features to the model, showing SHAP values across all samples. The color coding (red: high, blue: low) indicates feature importance. The plot effectively illustrates the distribution of each feature's impact on the model output [8]. A positive impact predicts risky drinking, while a negative impact suggests non-risky drinking.

Parents' education did not exhibit a clear pattern for students' alcohol consumption. Students with positive family relationships were less prone to risky consumption of alcohol, while those from families having fewer than 3 members and lacking family support were more likely to engage in risky drinking. Living apart from parents correlated with a reduced likelihood of risky alcohol behavior. Students with fathers in healthcare and education sectors tended to maintain safer alcohol intake, while those with fathers in services and other jobs did not ensure safe intake. Mothers' occupation at

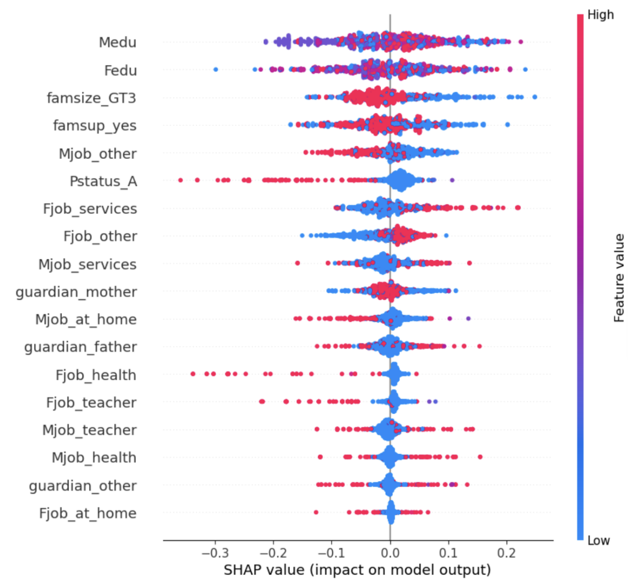

Figure 4: SHAP beeswarm plot

home indicated a lower likelihood of dangerous alcohol consumption, while other categories of mother's job showed mixed trends in students' alcohol consumption.

In our analysis, it is essential to underscore that the focus was on students as a demographic rather than on any specific geographical location. Consequently, the dataset primarily comprises numerical values, and the predictive factors are validated through algorithmic scrutiny, irrespective of locale. Given this context, it's reasonable to infer that similar results could be expected from any contextual dataset. Thus, the primary aim was to explore the underlying factors that exhibit a direct correlation with the predictive capabilities of the various models employed in this study.

5 LIMITATIONS

Several limitations of this study should be considered.

- Due to the predominantly categorical nature of the dataset, the encoding process resulted in a significant expansion of features. This complexity presented challenges in the manual selection of optimal features and the exploration of various feature subsets, impacting our ability to identify the best-performing model.
- One notable limitation of this research is the absence of a dedicated validation set for assessing the performance of the machine learning algorithms on an unseen dataset. This omission restricts our ability to evaluate the model's generalization capability beyond the training data.

6 CONCLUSION AND FUTURE WORK

In conclusion, this study introduces a machine learning framework tailored to predict risky alcohol consumption patterns among students. Through rigorous analysis, we have demonstrated the efficacy of machine learning in identifying students prone to hazardous drinking behaviors. Notably, our investigation highlights the superior performance of Random Forest and Decision Tree algorithms for our specific inquiry. Moreover, familial dynamics, particularly the quality of family relationships and parental cohabitation status, emerged as significant determinants of youths' alcohol consumption habits. We anticipate that this research will substantially contribute to mitigating risky alcohol consumption among students.

Looking ahead, we plan to expand the dataset to enhance model robustness and accuracy. Additionally, collaboration with domain experts, including psychologists and addiction specialists, will be pivotal in refining our approach and ensuring its relevance to real-world scenarios.

ACKNOWLEDGMENTS

The authors extend their sincerest appreciation to their project and research supervisor, Muhammad Shafayat Oshman, Lecturer in the Department of Electrical and Computer Engineering at North South University, Bangladesh. His unwavering support, precise guidance, and insightful advice have been instrumental throughout the experimental, research, and theoretical phases of this study, as well as in the preparation of this paper. Additionally, the authors gratefully acknowledge the Department of Electrical and Computer Engineering at North South University, Bangladesh, for their full funding provided for the publication of this research.

REFERENCES

[1] Mohammad H Afzali, Matthew Sunderland, Sherry Stewart, Benoit Masse, Jean Seguin, Nicola Newton, Maree Teesson, and Patricia Conrod. 2019. Machine-learning prediction of adolescent alcohol use: A cross-study, cross-cultural validation. *Addiction* 114, 4 (2019), 662–671.

[2] Aliaksandr Amialchuk, Onur Sapci, and Jon D Elhai. 2021. Applying machine learning methods to model social interactions in alcohol consumption among adolescents. *Addiction Research & Theory* 29, 5 (2021), 436–443.

[3] Abid Ali Awan. 2023. *An Introduction to SHAP Values and Machine Learning Interpretability.* https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability

[4] Ali Ebrahimi, Uffe Kock Wiil, Thomas Schmidt, Amin Naemi, Anette Søgaard Nielsen, Ghulam Mujtaba Shaikh, and Marjan Mansourvar. 2021. Predicting the risk of alcohol use disorder using machine learning: a systematic literature review. *IEEE Access* 9 (2021), 151697–151712.

[5] Ralph Hingson, Wenxing Zha, and Daniel Smyth. 2017. Magnitude and trends in heavy episodic drinking, alcohol-impaired driving, and alcohol-related mortality and overdose hospitalizations among emerging adults of college ages 18–24 in the United States, 1998–2014. *Journal of studies on alcohol and drugs* 78, 4 (2017), 540–548.

[6] Christopher M Jones, Heather B Clayton, Nicholas P Deputy, Douglas R Roehler, Jean Y Ko, Marissa B Esser, Kathryn A Brookmeyer, and Marci Feldman Hertz. 2020. Prescription opioid misuse and use of alcohol and other substances among high school students—Youth Risk Behavior Survey, United States, 2019. *MMWR supplements* 69, 1 (2020), 38.

[7] Suk-Young Kim, Taesung Park, Kwonyoung Kim, Jihoon Oh, Yoonjae Park, and Dai-Jin Kim. 2021. A deep learning algorithm to predict hazardous drinkers and the severity of alcohol-related problems using K-NHANES. *Frontiers in psychiatry* 12 (2021), 684406.

[8] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774.

[9] Grasiela Marcon, Flávia de Ávila Pereira, Aline Zimerman, Bruno Castro da Silva, Lisia von Diemen, Ives Cavalcante Passos, and Mariana Recamonde-Mendoza. 2021. Patterns of high-risk drinking among medical students: A web-based survey with machine learning. *Computers in Biology and Medicine* 136 (2021), 104747.

[10] Fabio Pagnotta and HM Amran. 2016. Using data mining to predict secondary school student alcohol consumption. *Department of Computer Science, University of Camerino* (2016), 1–9.

[11] Shamala Palaniappan, Norhamreeza A Hameed, Aida Mustapha, and Noor Azah Samsudin. 2017. Classification of alcohol consumption among secondary school students. *JOIV: International Journal on Informatics Visualization* 1, 4-2 (2017), 224–226.

[12] Thanthirige Lakshika Maduwanthi Ruberu, Emily A Kenyon, Karen A Hudson, Francesca Filbey, Sarah W Feldstein Ewing, Swati Biswas, and Pankaj K Choudhary. 2022. Joint risk prediction for hazardous use of alcohol, cannabis, and tobacco among adolescents: A preliminary study using statistical and machine learning. *Preventive Medicine Reports* 25 (2022), 101674.

[13] Rijad Sarić, Dejan Jokić, and Edhem Čustović. 2020. Identification of alcohol addicts among high school students using decision tree based algorithm. In *CMBEBIH 2019: Proceedings of the International Conference on Medical and Biological Engineering, 16 18 May 2019, Banja Luka, Bosnia and Herzegovina*. Springer, 459–467.

[14] Advait Singh, Vinayak Singh, Mahendra Kumar Gourisaria, and Ashish Sharma. 2022. Alcohol Consumption Rate Prediction using Machine Learning Algorithms. In *2022 OITS International Conference on Information Technology (OCIT)*. IEEE, 85–90.

[15] K Witkiewitz, RZ Litten, and L Leggio. 2019. Advances in the science and treatment of alcohol use disorder. *Science advances* 5, 9 (2019), eaax4043.