

Understanding Adolescent Alcohol Use: A Decision Tree Analysis of YRBSS Data

Isis Panellas¹ Catalina Canizares¹

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

Introduction

Substance use among adolescents in the United States represents a substantial public health challenge. According to the National Survey on Drug Use and Health, approximately 22.9% of individuals aged 12 to 17 have consumed alcohol at least once (SAMHSA, 2021). The ramifications of substance use during this developmental stage are extensive, with correlations to detrimental mental health outcomes (Esmaelzadeh et al., 2018; Kim, 2021; Meza et al., 2023), compromised academic performance (Bryant et al., 2003; Huỳnh et al., 2019), increased delinquency and criminal behavior (Childs et al., 2023), and higher instances of teen pregnancy (Russotti et al., 2023).

Given the prevalence and severe consequences of adolescent alcohol use, it is imperative to examine the factors that predispose youth to substance use, identifying those at elevated risk, and developing targeted preventative programs. This study used the 2021 Youth Risk Behavior Surveillance System (YRBSS) dataset to investigate the primary risk factors and their interplay. Our analysis seeks to understand patterns that could inform more efficacious preventive and intervention approaches, with the ultimate goal of preventing alcohol use among adolescents.

Methods

Data Source

- Utilized the 2021 Youth Risk Behavior Survey (YRBS) dataset.
- Weighted sample of 17,232 adolescents.
- Analyzed 55 variables, including alcohol and substance use, risky behaviors, dietary habits, hopelessness, suicidality, and physical activity.

Data Preprocessing

- Missing data assessed and determined to be missing at random (MAR).
- Multiple imputation methods used to handle missing values.

Resmapling methods

- 75% training set and 25% testing set to evaluate algorithm performance.
- Used five-fold cross-validation for hyperparameter tuning of the decision tree algorithm.
- Initial poor sensitivity led to the application of downsampling methods to balance classes. Algorithm retrained and tested using the balanced dataset.

Results

Table 1 presents the demographic distribution of current alcohol use from the 2021 YRBS data. Out of the sample, 4,045 individuals reported current alcohol use. A breakdown by gender indicates that 27% of females and 20% of males currently consume alcohol. American Indian/Alaska Natives lead in alcohol use prevalence, followed closely by White respondents and then by those identifying as Multiple-Hispanic. On the contrary, the Asian group reported the lowest alcohol consumption rates among all racial categories. In terms of sexual orientation, bisexual respondents reported the highest prevalence of alcohol use at 32%. This is followed by those who identified as “not sure”, while both homosexual and heterosexual groups presented similar rates of alcohol consumption

According to Figure 1

The study found a strong link between **vaping** and alcohol consumption among adolescents, with **87.5%** of those **currently vaping also using alcohol**.

Additional layers of complexity are introduced when considering other factors like **marijuana** and **pain medicine use, racial background, and hopelessness**. For example, among non-vapers, marijuana use and feelings of hopelessness are associated with **current alcohol consumption**.

Gender also plays a role, particularly in adolescents who do not vape but have used marijuana and belong to specific racial groups; **70.2% of males in this subset answered no to current use** use compared to **52.6% of females who currently consume it**.

Figure 1, Current Alcohol Use in Youth Decision Tree

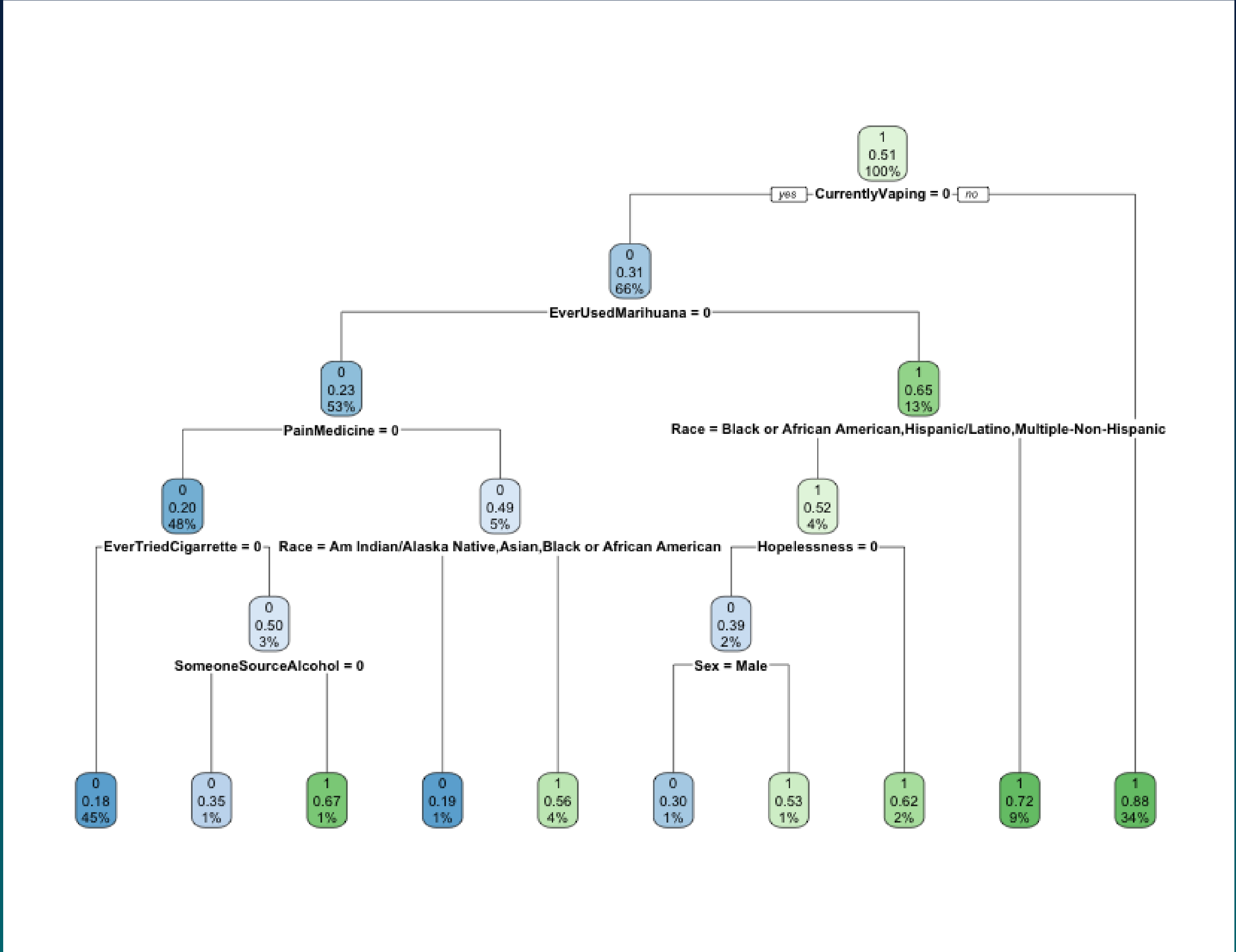


Table 1, Demographic Breakdown of Current Alcohol Use: Based on 2021 YRBS Data [n = 17,232]

| Current Alcohol Use | Yes, N = 4,045 ¹ | No, N = 13,187 ¹ |
|---------------------------|-----------------------------|-----------------------------|
| Sex | | |
| Female | 2,242 (27%) | 5,959 (73%) |
| Male | 1,728 (20%) | 7,055 (80%) |
| Race | | |
| White | 2,265 (27%) | 6,269 (73%) |
| Black or African American | 308 (15%) | 1,728 (85%) |
| Multiple-Hispanic | 669 (25%) | 1,986 (75%) |
| Hispanic/Latino | 339 (21%) | 1,280 (79%) |
| Multiple-Non-Hispanic | 235 (24%) | 725 (76%) |
| Asian | 91 (11%) | 738 (89%) |
| Am Indian/Alaska Native | 37 (32%) | 78 (68%) |
| Native Hawaiian/Other PI | 20 (23%) | 65 (77%) |
| SexOrientation | | |
| Heterosexual | 2,707 (22%) | 9,468 (78%) |
| Bisexual | 623 (32%) | 1,328 (68%) |
| Not sure | 152 (24%) | 480 (76%) |
| Gay or Lesbian | 114 (22%) | 408 (78%) |
| Grade | | |
| 9 | 706 (16%) | 3,850 (84%) |
| 10 | 874 (20%) | 3,477 (80%) |
| 11 | 1,100 (27%) | 3,050 (73%) |
| 12 | 1,316 (33%) | 2,696 (67%) |

Table 2 displays the performance metrics of the model on both training and testing datasets. On the training set, the model achieved a sensitivity of 0.80, indicating it correctly identified 80% of the positive cases. The specificity, representing the model's ability to correctly identify negative cases, stood at 0.77 for both training and testing sets. The overall accuracy, or the proportion of true results among the total number of cases examined, was 0.78 for the training set and 0.77 for the testing set. These metrics suggest that the model demonstrates consistent performance across both training and unseen data.

Table 2, Performance Metrics for the Classification Tree

| Dataset | Sensitivity | Specificity | Accuracy |
|----------|-------------|-------------|----------|
| Training | 0.80 | 0.77 | 0.78 |
| Testing | 0.78 | 0.77 | 0.77 |

Discussion

Our study clarify the risk factors contributing to alcohol use among adolescents. Results show vaping, marijuana, pain medication, and cigarette use as key contributors. These findings echo the notion of vaping as an entry point to other substance use, a stance supported by recent research (Boccio et al., 2022). Our analysis also highlights a prevalent trend of polysubstance use rather than isolated substance use, which is corroborated by data indicating a significant proportion of adolescents engaging in the concurrent use of alcohol and other substances (Goodwin et al., 2022).

Despite the insights provided, our study's cross-sectional design precludes us from inferring causation from the associations identified. Therefore, future research should implement longitudinal designs that can trace these risk factors over time, offering a clearer view of their causal relationships and the development of tailored preventive measures. Building on our study's foundation and addressing its limitations will not only enhance our understanding of adolescent alcohol use but also support the advancement of targeted intervention strategies.

References